

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF: Goan-Mook CHOI

GAU:

SERIAL NO: New Application

EXAMINER:

FILED: Herewith

FOR: MOBILE COMMUNICATION TERMINAL HAVING VOICE RECOGNITION FUNCTION, AND
PHONEME MODELING METHOD AND VOICE RECOGNITION METHOD FOR THE SAME

REQUEST FOR PRIORITY

COMMISSIONER FOR PATENTS
ALEXANDRIA, VIRGINIA 22313

SIR:

- ☐ Full benefit of the filing date of U.S. Application Serial Number , filed , is claimed pursuant to the provisions of 35 U.S.C. §120.
- ☐ Full benefit of the filing date(s) of U.S. Provisional Application(s) is claimed pursuant to the provisions of 35 U.S.C. §119(e): Application No. Date Filed
- ☒ Applicants claim any right to priority from any earlier filed applications to which they may be entitled pursuant to the provisions of 35 U.S.C. §119, as noted below.

In the matter of the above-identified application for patent, notice is hereby given that the applicants claim as priority:

<u>COUNTRY</u>	<u>APPLICATION NUMBER</u>	<u>MONTH/DAY/YEAR</u>
Korea	10-2003-0069219	October 6, 2003

Certified copies of the corresponding Convention Application(s)

- ☒ are submitted herewith
- ☐ will be submitted prior to payment of the Final Fee
- ☐ were filed in prior application Serial No. filed
- ☐ were submitted to the International Bureau in PCT Application Number
Receipt of the certified copies by the International Bureau in a timely manner under PCT Rule 17.1(a) has been acknowledged as evidenced by the attached PCT/IB/304.
- ☐ (A) Application Serial No.(s) were filed in prior application Serial No. filed ; and
- ☐ (B) Application Serial No.(s)
- ☐ are submitted herewith
- ☐ will be submitted prior to payment of the Final Fee

Respectfully Submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.



Marvin J. Spivak

Registration No. 24,913

Customer Number

22850

Tel. (703) 413-3000
Fax. (703) 413-2220
(OSMMN 05/03)

C. Irvin McClelland
Registration Number 21,124



별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto is a true copy from the records of the Korean Intellectual Property Office.

출원 번호 : 10-2003-0069219
Application Number

출원 년 월 일 : 2003년 10월 06일
Date of Application OCT 06, 2003

출원인 : 주식회사 팬택앤큐리텔
Applicant(s) Curitel Communications, Inc.



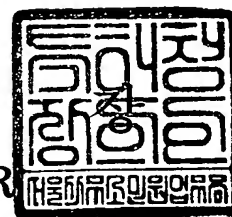
2003 년 11 월 11 일

특

허

청

COMMISSIONER



【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【제출일자】	2003.10.06
【발명의 명칭】	음성인식 기능을 가진 이동 통신 단말기, 및 이를 위한 음소 모델링 방법 및 음성 인식 방법
【발명의 영문명칭】	Mobile Communication Terminal with Voice Recognition function, Phoneme Modeling Method and Voice Recognition Method for the same
【출원인】	
【명칭】	(주)팬택앤큐리텔
【출원인코드】	1-2001-021691-6
【대리인】	
【명칭】	특허법인 엘엔케이
【대리인코드】	9-2000-100002-5
【지정된변리사】	변리사 이헌수
【포괄위임등록번호】	2003-024662-6
【발명자】	
【성명의 국문표기】	최관묵
【성명의 영문표기】	CHOI, GOAN MOOK
【주민등록번호】	740319-1063826
【우편번호】	137-042
【주소】	서울특별시 서초구 반포2동 대우아파트 101-1101
【국적】	KR
【심사청구】	청구
【취지】	특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인 특허법인 엘엔케이 (인)
【수수료】	
【기본출원료】	15 면 29,000 원
【가산출원료】	0 면 0 원
【우선권주장료】	0 건 0 원
【심사청구료】	3 항 205,000 원
【합계】	234,000 원

【요약서】**【요약】**

본 발명은 음성인식 기능을 가진 이동 통신 단말기에 관한 것으로서, 특히 이동 통신 단말기에서 음성인식률을 향상시키기 위한 음소 모델링 방법과 그에 기초한 음성 인식방법에 관한 것이다. 본 발명에 따른 이동 통신 단말기는 음성을 입력받기 위한 음성입력기와; 사용자가 입력한 음성에 의해 생성된 각 음소별 특징벡터의 기준 모델을 저장하고 있는 저장부와; 입력 음성의 음소들을 구분하여 각 음소별 특징벡터를 추출하고 추출된 음소별 특징벡터를 상기 음소별 기준 모델들과 패턴매칭 수행하여 음성 인식하는 제어부;를 포함함을 특징으로 한다.

【대표도】

도 2

【색인어】

화자종속, 반복 훈련, 음성 인식.

【명세서】**【발명의 명칭】**

음성인식 기능을 가진 이동 통신 단말기, 및 이를 위한 음소 모델링 방법 및 음성 인식 방법 {Mobile Communication Terminal with Voice Recognition function, Phoneme Modeling Method and Voice Recognition Method for the same}

【도면의 간단한 설명】

도 1은 본 발명의 실시예에 따른 이동 통신 단말기의 블록구성도.

도 2는 본 발명의 실시예에 따른 음소 모델링 절차 흐름도.

도 3은 본 발명의 실시예에 따른 음소 모델링에 기초하여 음성인식을 하는 과정을 설명하기 위한 흐름도.

【발명의 상세한 설명】**【발명의 목적】****【발명이 속하는 기술분야 및 그 분야의 종래기술】**

<4> 본 발명은 이동 통신 단말기의 음성 인식에 관한 것으로, 특히 음성 인식을 위한 음소 모델링 방법과 그에 기초한 음성 인식방법 및 이를 채용한 이동 통신 단말기에 관한 것이다.

<5> 음성 인식 시스템이란 입력 수단의 하나로서 사용자의 음성을 인식하여 그에 해당하는 작업을 수행하는 시스템이다. 이러한 음성 인식 시스템은 입력된 음성의 특징을 추출하여 기존에 훈련된 기준 음성 모델들과 패턴매칭을 수행하여 음성을 구별하는 시스템으로, 상기 기준 음성 모델을 훈련하는 과정은 그 횟수가 많아질수록 더 일반적인 기준 음성 모델을 구할 수 있다.

<6> 상술한 음성 인식 시스템의 하나로서 화자 종속형 음성 인식 시스템을 들 수 있다. 이동 통신 단말기의 경우 사용자가 1인이므로 사용자의 음성을 이용해서 DB를 구축하는 것이 적절하다. 이러한 이유로 이동 통신 단말기에서는 화자 종속형 음성 인식 시스템을 주로 채용하고 있다. 그러나 이동 통신 단말기의 화자 종속형 음성 인식 시스템은, 예를 들어, "우리집"과 같은 단어의 반복 입력을 통해 그 입력단어에 대한 기준 음성을 모델링하기 때문에, 사용자 입장에서 보면 음성 다이얼링 혹은 단말기 제어를 위해 필요한 단어들(예를 들면, 우리집, 회사, 시댁,...)을 매번 반복 입력하여 기준 모델을 생성해야 하는 불편함이 있다.

<7> 또한 종래 이동 통신 단말기의 음성 인식 시스템은 그 특성상 반복 훈련을 통해서 음성 인식률을 높일 수 있도록 설계되어 있으나, 이동 통신 단말기에서 채용하고 있는 음성 인식 시스템은 프로그램적으로 단어의 음성 입력 횟수를 미리 제한하고 있거나 미리 기준 모델을 DB화해 놓았기 때문에 음성 인식률을 향상시키는데 한계가 있다.

【발명이 이루고자 하는 기술적 과제】

<8> 따라서 본 발명의 목적은 음성 인식률을 향상시킬 수 있는 음소 모델링 방법 및 음성 인식 방법을 제공하기 위한 것이며,

<9> 더 나아가 본 발명의 또 다른 목적은 음성 인식률을 향상시킬 수 있는 음성 인식 방법을 채용한 이동 통신 단말기를 제공하기 위한 것이다.

【발명의 구성 및 작용】

<10> 상기 목적을 달성하기 위한 본 발명의 실시예에 따른 이동 통신 단말기는,

<11> 음성을 입력받기 위한 음성입력기와;

- <12> 사용자가 입력한 음성에 의해 생성된 각 음소별 특징벡터의 기준 모델을 저장하고 있는 저장부와;
- <13> 입력 음성의 음소들을 구분하여 각 음소별 특징벡터를 추출하고 추출된 음소별 특징벡터를 상기 음소별 기준 모델들과 패턴매칭 수행하여 음성 인식하는 제어부;를 포함함을 특징으로 한다.
- <14> 아울러 본 발명에 따른 이동 통신 단말기의 음소 모델링 방법은,
- <15> 표시 문자에 대한 음성을 입력받는 단계와;
- <16> 입력된 음성을 표시 문자의 음소들로 구획하는 단계와;
- <17> 구획된 각 음소들의 특징벡터를 추출하는 단계와;
- <18> 추출된 각 음소들의 특징벡터를 음소별로 분류되어 있는 이전의 해당 음소별 특징벡터와 연산하여 각 음소별 기준 모델을 생성 저장하는 단계;를 포함함을 특징으로 한다.
- <19> 그리고 본 발명에 따른 이동 통신 단말기의 음성인식 방법은,
- <20> 사용자가 입력한 음성에 의해 생성된 각 음소별 특징벡터의 기준 모델을 생성하여 저장하는 단계와;
- <21> 음성을 입력받는 단계와;
- <22> 상기 입력된 음성의 음소들을 구분하여 각 음소별 특징벡터를 추출하는 단계와;
- <23> 상기 추출된 음소별 특징벡터를 저장된 상기 음소별 기준 모델들과 패턴매칭 수행하여 음성을 인식하는 단계;를 포함함을 특징으로 한다.

- <24> 상술한 방법을 통해 본 발명은 한글 자모 각각에 대한 기준 모델을 생성하여 놓을 수 있기 때문에, 사용자가 표시 문자에 대한 음성을 입력하는 것만으로 각 음소별 기준 모델이 지속적으로 갱신되기 때문에 결과적으로는 음성 인식률을 향상시킬 수 있는 효과를 누릴 수 있다.
- <25> 또한, 모든 한글 어휘에 대한 음성인식이 가능함으로서 사용자는 음성 인식을 위해 필요한 음성들을 반복 입력하여야만 하는 불편함으로부터 해방될 수 있는 이점도 있다.
- <26> 이하 본 발명의 바람직한 실시예를 첨부 도면을 참조하여 상세히 설명하기로 한다. 본 발명을 설명함에 있어 관련된 공지 기능 혹은 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그에 대한 상세한 설명은 생략하기로 한다.
- <27> 도 1은 본 발명의 실시예에 따른 이동 통신 단말기의 블록구성도를 도시한 것으로, 보다 구체적으로는 카메라 폰의 블록구성을 도시한 것이다.
- <28> 도 1을 참조해 보면, RF(Radio Frequency) 모듈(module)(100)은 안테나(ANT)를 통해 기지국으로부터 수신되는 RF 신호를 복조하여 베이스밴드 처리부(102)에 제공하고, 베이스밴드 처리부(102)가 제공하는 신호를 RF 신호로 변조하여 안테나(ANT)를 통해 기지국으로 송신한다.
- <29> 베이스밴드(BB) 처리부(102)는 RF 모듈(100)의 출력신호를 다운 컨버전한 후에 디지털 신호로 변환하여 제어부(104)에 제공하고, 제어부(104)로부터 제공되는 신호를 아날로그 신호로 변환한 후에 업 컨버전하여 RF모듈(100)에 제공한다.
- <30> 제어부(104)는 후술할 메모리부(106)에 저장된 제어프로그램 데이터에 기초하여 도시된 카메라 폰의 동작을 전반적으로 제어한다. 예를 들면 상기 제어부(104)는 도 2와 도 3에 도시한 바와 같은 수순에 따라 각 음소별 기준 모델을 생성하여 저장하고, 사용자에 의해 입력된

음성을 구성하는 음소들의 특징을 추출하여 상기 음소별 기준 모델과 패턴매칭을 수행함으로써 음성을 인식한다.

- <31> 메모리부(106)에는 적어도 카메라 폰의 동작을 제어하기 위한 제어 프로그램 데이터와 후술할 카메라(110)에 의해 촬상된 영상 데이터 및 본 발명의 실시예에 따른 음소별 기준 특징 벡터들(기준 모델이라고도 함)이 저장된다.
- <32> 키패드(key pad)(108)는 문자를 입력하기 위한 사용자 인터페이스로서 이미 공지된 바와 같이 4*3의 문자키와 다수의 기능키를 포함한다. 이러한 키패드(108)를 문자입력기로 명할 수도 있다.
- <33> 한편 카메라(110)는 피사체를 촬상하여 출력하여 주는 역할을 수행하며, 영상신호 처리부(112)는 카메라(110)로부터 출력되는 촬상신호를 신호처리하여 한 프레임의 영상을 생성하고 이를 출력하여 주는 역할을 수행한다.
- <34> 음성입력부(114)는 마이크를 통해 입력된 음성신호를 증폭하여 디지털 데이터로 변환하고 이를 음성인식을 위해 필요한 신호로 처리하여 제어부(104)로 출력한다. 이러한 음성입력부(114) 역시 음성입력기로 명할 수 있을 것이다.
- <35> 표시부(116)는 제어부(104)의 제어에 따라 문자 혹은 촬상된 영상데이터를 표시하여 주는 역할을 수행한다.
- <36> 이하 상술한 구성을 가지는 카메라 폰에서 사용자 음성에 대한 음소의 특징들을 모델링하여 음성 인식에 이용하는 과정을 설명하기로 한다.
- <37> 우선 본 발명의 실시예에 따른 음소 모델링 방법은 카메라 폰 사용자가 발음하는 문자에 대한 음성을 최소 단위인 음소로 구분하고 각 음소별 기준 모델을 생성하여 데이터베이스화하

는 단계와, 입력되는 음성의 음소들을 구분하여 각 음소별 특징벡터를 추출하고 그 추출된 음소별 특징벡터를 데이터베이스화된 음소별 기준 모델들과 패턴매칭 수행하여 음성 인식하는 단계로 구분할 수 있다.

<38> 각 음소별 기준 모델을 생성하여 데이터베이스화하는 단계는 도 2에 의해 설명될 수 있으며, 음성 인식하는 단계는 도 3을 참조하여 설명될 수 있다. 참고적으로 본 발명에서의 음소는 문자를 구성하는 최소 단위 즉, 자음과 모음, 음성을 구성하는 자음과 모음을 지칭하는 것으로 가정한다.

<39> 도 2를 우선 참조하면, 카메라 폰 사용자가 음성 입력만으로 카메라 폰을 제어하기 위해서는 우선적으로 자음과 모음에 대한 음소별 기준 모델을 생성하여야 한다. 이러한 음소별 기준 모델 생성을 위해 사용자가 음소 특징추출모드의 수행을 명하면, 제어부(104)는 이를 200단계에서 감지하고 사용자에게 문자의 입력(혹은 선택)을 요구한다. 이때의 문자는 사용자가 키패드(108)를 통해 입력하는 문자일 수 있으며, 경우에 따라서는 수신된 문자 메시지일 수도 있다. 중요한 것은 입력 혹은 선택된 문자가 표시부(116)상에 표시된후, 이 표시 문자를 사용자가 음성으로 입력함으로써 입력된 각각의 음성을 구성하는 음소들의 기준 모델이 생성된다는 것이다.

<40> 만약 사용자가 훈련시키고자 하는 문자(예를 들면 "가")를 입력(210단계)하면, 제어부(104)는 입력문자에 대한 음성 입력을 요구한다. 이에 사용자 자신이 입력한 문자에 대해 발음하면, 문자에 대한 음성이 음성입력부(114)를 통해 입력(220단계)된다.

<41> 제어부(104)는 음성입력부(114)를 통해 입력문자에 대한 음성 입력이 있으면 입력음성의 음소를 구분("ㄱ"과 "ㅏ")하고, 구분된 음소의 특징벡터를 추출(230단계)한다. 그리고 제어부(104)는 240단계로 진행하여 상기 230단계에서 추출된 음소의 특징벡터를 음소별 기준 특징벡

터로 설정하여 저장한다. 230단계에서 추출된 음소의 특징벡터가 기준 특징벡터로 설정되는 이유는 최초의 문자 입력이 이루어졌다고 가정할 수 있기 때문이다.

<42> 만약 사용자가 "나"라는 새로운 문자를 입력(210단계)하고 그에 대해 음성 입력하면, 제어부(104)는 230단계를 수행함으로써, "ㄴ"음소에 대해서는 두 번의 특징벡터 추출이 이루어진 결과를 갖게 된다. 따라서 음소 "ㄴ"에 대해 얻어진 두 개의 특징벡터들의 평균값을 산출하면 이를 기준 특징 벡터로 설정할 수 있다. 예시한 바에 의하면 "ㄱ"과 "ㄴ", "ㄴ"에 대한 기준모델이 생성된 결과를 얻게 된다.

<43> 즉, 본 발명은 사용자 자신이 입력한 혹은 선택한 문자들에 대해 각각 음성 입력하면, 각 음성을 구성하는 음소들에 대한 특징벡터들이 추출되고, 추출된 각 음소의 특징벡터들은 이전에 저장된 해당 음소의 기준 특징벡터와 연산되어 새로운 기준 특징벡터로 생성됨으로서, 데이터베이스화된 각 음소별 기준 모델들이 반복 훈련을 통해 지속적으로 갱신될 수 있는 것이며, 자음과 모음 모두에 대한 기준 모델을 생성할 수 있게 되는 것이다.

<44> 상술한 바와 같은 방법을 통해 생성된 음소별 기준모델을 이용하여 음성 인식이 이루어지는 과정을 이하 도 3을 참조하여 설명하면,

<45> 우선 음성인식모드에서 제어부(104)는 음성 입력부(114)를 통해 음성이 입력되는가를 검사(300단계)한다. 만약 우리 집을 호출하기 위한 음성정보로서 "우리 집"이 음성 입력되었다면 제어부(104)는 입력된 음성의 음소들을 구분하고 각 음소별 특징벡터를 추출(310단계)한다. 이어서 제어부(104)는 추출된 음소별 특징벡터를 메모리부(106)에 저장된 음소별 기준 모델들과 패턴매칭 수행(320단계)한다. 이러한 패턴매칭에는 HMM(Hidden Markov Model) 알고리즘을 적용할 수 있다.

<46> 한편 제어부(104)는 패턴매칭 수행 결과 310단계에서 추출된 음소의 특징벡터와 기준 모델의 거리가 최소인 음소들을 추출하여 조합하는 방식을 통해 음성을 인식(330단계)한다. 그리고 인식된 결과를 처리하는데, 그러한 인식 결과의 예가 바로 자동 다이얼링을 수행하는 것이다. 자동 다이얼링이 이루어지도록 하기 위해서는 물론 사전에 우리집에 대한 전화번호가 우리 집:02-888-8888과 같이 미리 등록되어 있어야만 한다.

<47> 실시예로서 설명한 바와 같이, 본 발명은 한글 어휘 모두에 대한 음성을 인식할 수 있도록 미리 음소별 기준모델을 사용자가 생성하여 놓았기 때문에, 사용자는 예시한 바와 같이 사전에 "우리집"이라는 단어의 음성을 반복 입력하지 않고서도 우리 집이라는 음성 입력만으로 우리 집을 호출할 수 있게 되는 것이다.

【발명의 효과】

<48> 상술한 바와 같이 본 발명은 표시 문자에 대한 사용자 음성을 입력받아 그 음성을 구성하는 음소들의 기준 모델을 지속적으로 갱신하기 때문에, 음성 인식률을 향상시키게 되는 효과를 얻을 수 있다. 또한, 인식하고자 하는 단어에 대한 음성의 반복 훈련 없이도 추후 그 단어를 음성 인식할 수 있는 장점이 있다. 이는 곧 모든 한글 어휘에 대한 음성인식이 가능하다는 것을 의미한다.

<49> 한편 본 발명은 도면에 도시된 실시예들을 참고로 설명되었으나 이는 예시적인 것에 불과하며, 당해 기술분야에 통상의 지식을 지닌 자라면 이로부터 다양한 변형 및 균등한 타실시예가 가능하다는 점을 이해할 것이다. 따라서 본 발명의 진정한 기술적 보호범위는 첨부된 특허 청구범위에 의해서만 정해져야 할 것이다.

【특허청구범위】**【청구항 1】**

음성을 입력받기 위한 음성입력기와;

사용자가 입력한 음성에 의해 생성된 각 음소별 특징벡터의 기준 모델을 저장하고 있는 저장부와;

입력 음성의 음소들을 구분하여 각 음소별 특징벡터를 추출하고 추출된 음소별 특징벡터를 상기 음소별 기준 모델들과 패턴매칭 수행하여 음성 인식하는 제어부;를 포함함을 특징으로 하는 이동 통신 단말기.

【청구항 2】

표시 문자에 대한 음성을 입력받는 단계와;

입력된 음성을 표시 문자의 음소들로 구획하는 단계와;

구획된 각 음소들의 특징벡터를 추출하는 단계와;

추출된 각 음소들의 특징벡터를 음소별로 분류되어 있는 이전의 해당 음소별 특징벡터와 연산하여 각 음소별 기준 모델을 생성 저장하는 단계;

를 포함함을 특징으로 하는 이동 통신 단말기의 음소 모델링 방법.

【청구항 3】

사용자가 입력한 음성에 의해 생성된 각 음소별 특징벡터의 기준 모델을 생성하여 저장하는 단계와;

음성을 입력받는 단계와;

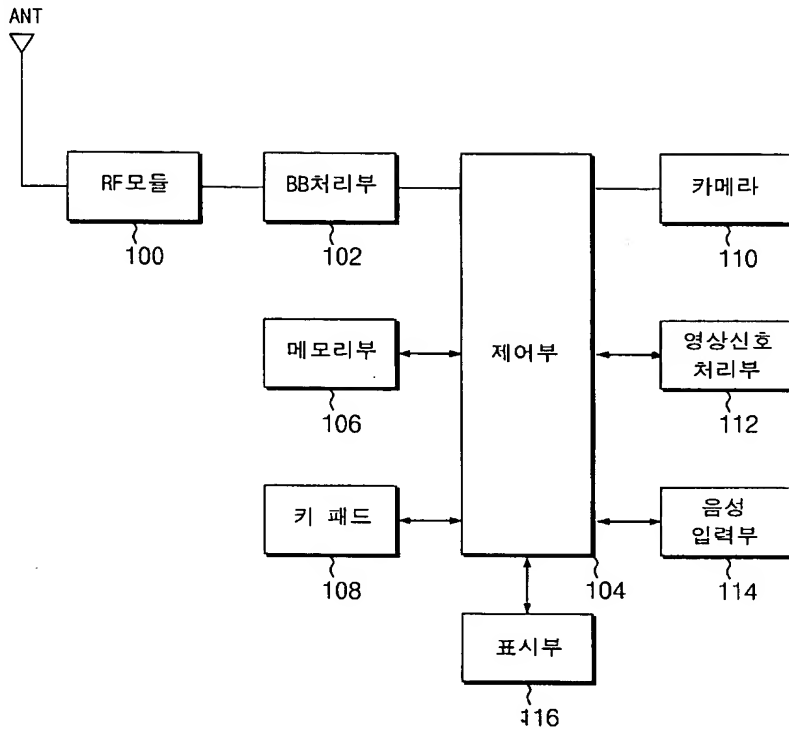
상기 입력된 음성의 음소들을 구분하여 각 음소별 특징벡터를 추출하는 단계와;

상기 추출된 음소별 특징벡터를 저장된 상기 음소별 기준 모델들과 패턴매칭 수행하여
음성을 인식하는 단계;

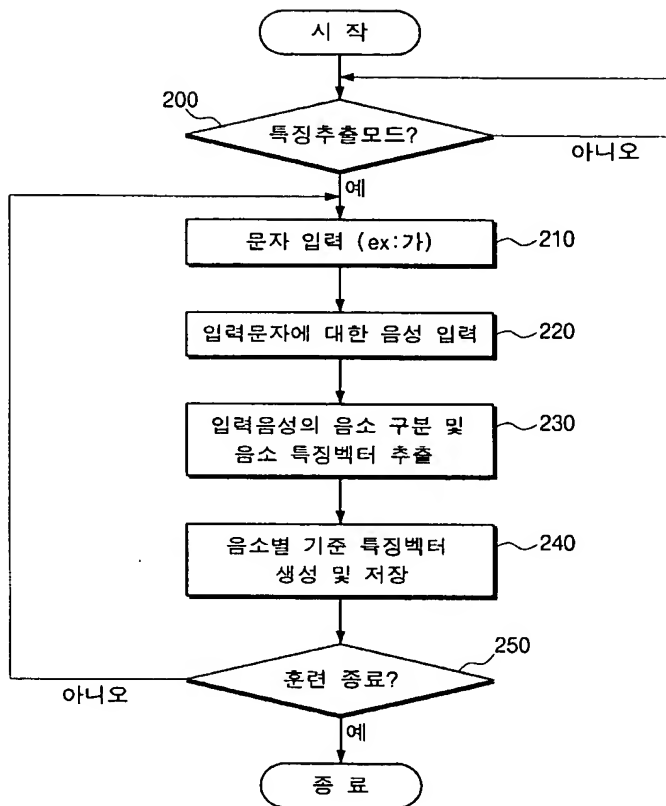
를 포함함을 특징으로 하는 이동 통신 단말기의 음성 인식 방법.

【도면】

【도 1】



【도 2】



【도 3】

